**CAPSTONE PROJECT**

**PROJECT SYNOPSIS**

**GROUP-IV**

**Members: Mentor:**

1.ANIRUDH DYAGA VIBHA

2.YASHWANTHRAJ GUDISE

3.SAI KUMAR GANDLA

4.SITAROJ ANILKUMAR

5.BOMMANA SRAVAN

6.ROHITH KANDULA

**PROBLEM STATEMENT:**

To predict the authenticity of a job-post given by the companies after observing their choice of words in the Job title/ Job description.

**ABSTRACT:**

Since the world is struck by once in a century kind of pandemic the graph of economy is suffering. If the economy graph is experiencing a recession in such scenario, then the pace of joblessness will in general ascent because of the absence of occupations. Amidst of such high unemployment rates there exists another problem, “Fake jobs”. Fake job postings are dangerous as most of the unemployed would be desperate to apply for all the jobs they come across. While applying for such jobs the applicants will eventually give in their valuable data which includes date of birth, contact details and education details. In this project, we showcase the patterns and differences between a real job posting and a fake job posting.

**METHODOLOGIES:**

* Firstly, pre-process the data by using exploratory data analysis (EDA) techniques
* Then the various Natural language processing (NLP) methodologies such as Bag of Words, text processing and word embeddings are used to classify the words used in the job.
* The base model will be built by using supervised classification algorithms (SLC) such as naïve bayes and logistic regression to predict the binary outcome that is either the job is fake or not.
* Lastly fine tune the model with advanced techniques like feature extraction, balancing the data and ensemble methods.

**TIMELINE CHART (Weekly Plan)**

|  |  |
| --- | --- |
| **WEEK** | **PLAN OF ACTION** |
| Week-1 | Group creation and domain selection. |
| Week-2 | Collection of datasets and interaction with the mentor. |
| Week-3 | Finalization of dataset and synopsis preparation. |
| Week-4 & 5 | Data preprocessing, Missing value treatment, EDA analysis (Univariate, Bivariate and Multivariate analysis), Outlier treatment. |
| Week-6 | Getting business insights and Basic model building and Interim report preparation. |
| Week-7 | Interim report presentation. |
| Week-8 & 9 | Feature engineering & feature extraction  Hyper parameters tuning |
| Week-10 | Comparison and selection of model  Results & Discussion. |
| Week-11 | Conclusions & Recommendations |
| Week-12 | Final Report preparation. |
| Week -13 | Final Report submission. |
| Week -14 | Final report presentation. |

**References:**

Dataset collection repository-

[[Real or Fake] Fake JobPosting Prediction | Kaggle](https://www.kaggle.com/shivamb/real-or-fake-fake-jobposting-prediction)

Unemployment rate in world –

[Unemployment By Country 2021 (worldpopulationreview.com)](https://worldpopulationreview.com/country-rankings/unemployment-by-country)